DOCUMENT RESUME

ED 273 918 CG 019 383

AUTHOR Haaga, John
TITLE The Current State of Family Planning Program Evaluation.
INSTITUTION Rand Corp., Santa Monica, Calif.
SPONS AGENCY Agency for International Development (IDCA), Washington, DC.
REPORT NO Rand-N-2337-AID
PUB DATE Jul 85
GRANT AID/OTR-G-1822
NOTE 27p.
AVAILABLE FROM The Rand Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138 ($4.00; 25% discount on 25 or more).
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Birth Rate; *Developing Nations; Evaluation Criteria; *Evaluation Methods; *Family Planning; *Program Evaluation; *Research Problems

ABSTRACT

This report provides an overview of methods of evaluating the impact of family planning programs in developing countries, with a discussion of their strengths and weaknesses. The focus is on estimation of program impact on fertility rates, though other potential outcome measures, such as maternal and child health, are also discussed. The first section sets the context of recent debates over the effectiveness of family planning programs. The following sections deal with methods of analysis and data requirements for simple standardization of fertility rates for before-and-after comparisons, quasi-experimental designs and multivariate analyses of survey data, and components models for simulating impacts. The tradeoff in designing evaluations is between the difficulty and expense of the data collection and analysis and the reliability of inferences about program impact, since the more reliable and plausible methods demand data not routinely available in countries. The last section discusses the need for estimates of marginal program effectiveness as well as average effectiveness, for resource allocation decisions. (Author/NB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

This research was supported by the U.S. Agency for International Development under Grant No. AID/OTR-G-1822.

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

PREFACE

This Note is a revised version of a paper written for a tutorial on "Evaluation of Human Resource Programs" at the Rand Graduate Institute in 1981. It provides an overview of methods of evaluating the impact of family planning programs in developing countries, with a discussion of their strengths and limitations. The paper was shortened, and some sections dealing with unpublished evaluations were dropped for the present version; but the text was not extensively revised to take account of studies published since 1981.

The Note is issued by the Family in Economic Development Center of The Rand Corporation with the support of the U.S. Agency for International Development under Grant AID/OTR-G-1822. It should be of interest to the staff of the Agency and to other evaluators of family planning programs.

SUMMARY

This Note reviews the major methods of evaluating the impact of family planning programs in developing countries. The focus is on estimation of program impact on fertility rates, though other potential outcome measures, such as maternal and child health, are also discussed. The first section sets the context of recent debates over the effectiveness of family planning programs. The following sections deal with methods of analysis and data requirements for simple standardization of fertility rates for before-and-after comparisons, quasi-experimental designs and multivariate analyses of survey data, and components models for simulating impacts. The tradeoff in designing evaluations is between the difficulty and expense of the data collection and analysis and the reliability of inferences about program impact, since the more reliable and plausible methods demand data not routinely available in countries. The last section discusses the need for estimates of marginal program effectiveness as well as average effectiveness, for resource allocation decisions.

ACKNOWLEDGMENTS

The author would like to thank Dennis De Tray, William Butz, and Linda Waite for very helpful comments on an earlier version of this Note, and Nora Wolverton and Barbara Eubank-Thurston for preparing the manuscript.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

Section

INTRODUCTION 1
METHODS OF EVALUATING DEMOGRAPHIC IMPACT 4
MARGINAL EFFECTIVENESS 14
CONCLUSIONS 17

REFERENCES 19

THE CURRENT STATE OF FAMILY PLANNING PROGRAM EVALUATION

INTRODUCTION

This Note reviews the major methods of evaluating family planning programs, and illustrates problems in design and analysis of evaluation studies with examples that deal with national family planning programs in developing countries. Its focus, like that of the vast majority of program evaluations in the published literature, is on evaluations of the impact on fertility rates; the necessary preliminary steps of process evaluation--making sure that the program actually came about--and other forms of impact, notably on maternal and child health, are also considered. The first section sets the context of recent debates over the effectiveness of family planning programs. The following sections deal with methods of analysis and data requirements for simple standardization of fertility rates for trend analysis, quasi-experimental designs and multivariate analyses of survey data, and components models for simulating impacts. The tradeoff in designing evaluations is between the difficulty and expense of the data collection and analysis and the reliability of inferences about program impact, since the more reliable and plausible methods demand data that are not routinely available in developing countries. The last section discusses the need for some estimates of marginal program effectiveness as well as average effectiveness, for resource allocation decisions.

The Family Planning Debate

The very notion of family planning programs as they were conceived in the 1960s was daring. The international donor community was to provide the technology, materials, and advice, while the local governments were to deliver services, preferably outside the usual institutional structures of Health Ministries. Besides filling a presumed backlog of demand on the part of those who wanted to control their fertility but somehow lacked the means, the programs would generate new demand. Success would be measured by the numbers of "acceptors," and ultimately, by declines in fertility rates more rapid than predicted by theories of demographic modernization.

This could be viewed as one more manifestation of the social policy euphoria of that decade. There was little reason to believe that the torpid governments of the Third World, what Gunnar Myrdal aptly described as "the Soft States," could influence behavior in the most custom-bound, the most private aspects of human life. Far better established governments, like France in the 1930s, had sought to influence fertility behavior in the opposite direction, without notable success. Yet by 1980, a review of the efforts of the previous two decades contained the following claim:

> Family planning programs have significant, independent effects [on fertility decline], certainly in developing countries with favorable social settings and under certain conditions in less favorable settings as well . . . Moreover, the longer the program and the clearer its demographic intent, the greater its effect. [Berelson, Mauldin, and Segal, p. 74].

Against such claims, many argued that family planning programs at most facilitated fertility declines that were taking place anyway, due to basic changes in social relations and economic incentives. Couples have always had means for regulating fertility. When society changes in profound ways, resulting in a desire on the part of many couples to have fewer children than their own parents considered right and necessary, then, and not before, fertility declines. If modern methods of contraception are provided free of charge through a public program rather than through private channels, this is a subsidy to the couple, to be valued at the value they place on the money they have saved. But, in this view, the family planning programs did not cause the fertility decline in the sense that the decline would not have happened without it. This view is associated with the slogan that prevailed at the 1974 World Population Conference at Bucharest: "Development is the best contraceptive."[1]

---

[1]At the 1974 Conference, the notion that family planning programs by themselves have "significant, independent effects" was promoted especially by the American delegation. The opposing policy, that population problems would take care of themselves once economic development was achieved, was associated with the People's Republic of China. At the 1984 World Population Conference in Mexico City, the two delegations had switched sides.

One goal of program evaluations is to break through this impasse. In essence, impact evaluations are carried out to test the counterfactual hypothesis that a fertility decline would not have happened or would have been slower, had the program not existed. As one manual on family planning program evaluation puts it, "Questions that involve measurement of nonevents are inherently speculative. Yet they are not only commonplace but often the essence of program evaluation." [Nortman et al., 1978, p. 35]. The following sections of this paper review some of the major methods that have been devised to give some empirical backing to this "inherently speculative" enterprise.

Criteria for Evaluation

In the literature on the evaluation of family planning programs, there is virtual unanimity on one point: The main criterion on which the programs are to be judged is the decline in fertility rates they should cause. Intermediate measures, such as number of acceptors, may be used for the setting of targets, but claims for the success of a program usually refer to a coincident decline in a relevant fertility rate. It is worth pointing out, as Kingsley Davis did in an influential article more than a decade ago, that there is no simple connection between "family planning" and fertility reduction, let alone "population control": "The 'planning' in family planning is that of each separate couple." [Davis, p. 732]. If the interests of the couple, or what they perceive to be their interests, do not coincide with the plans of the State for lower birth rates, then simple distribution of contraceptives is not a population policy. Jack Reynolds pointed out in a perceptive review that it is not at all clear that "births averted" are the only successes--births that occur just when desired may be successes, as well [Reynolds, 1972].

Family planning may still be a boon to its recipients, independent of an effect on the fertility rates. The traditional means by which couples have regulated their fertility have always included abortion. Providing modern contraceptives to women who otherwise would rely on abortions may lessen the need to resort to an often dangerous operation. Another traditional means of spacing births is abstinence from sexual

relations for long periods after a birth--so that, even if modern contraceptive methods did no more than substitute for this traditional method, the increase in personal welfare of couples might be judged important. Where women desire or have to do work that separates them from infants, modern contraceptives may substitute for the birth-spacing effect of prolonged breastfeeding. The modern methods provided by programs, when used correctly, are more effective than are traditional methods of birth control. An increase in the ability of people to control their own fertility--their own lives--must be of some value to them, reducing uncertainty about economic or physical hardship.

In the early stages of many of the family-planning national movements, when voluntary agencies were organizing services despite official indifference or even hostility, family planning was usually touted as a maternal and child health service. Their demographic purpose was not exactly hidden but was in many places certainly de-emphasized, as is still true in many African and Asian countries where ethnic or racial balance is a delicate subject. Since the 1974 Bucharest Conference, it has increasingly been argued that family planning services should be integrated into health services, rather than stand apart as instruments of population policy, in large part to reflect this health care orientation. Given this important and often explicit goal, it is remarkable how rarely evaluators have attempted to estimate the effects of family planning programs on *mortality*, rather than fertility rates. [Wolfson, 1972].

METHODS OF EVALUATING DEMOGRAPHIC IMPACT

In the next section of this paper, the major methods available to evaluate demographic impacts of family planning programs are reviewed, progressing from methods that involve little more than before-and-after comparisons to methods that yield more reliable inferences, though with greater cost and complexity.

Standardization

The very crudest method of ascertaining the impact of a family planning program on fertility rates is simply to compare rates before and after implementation of a program. Such a naive approach ignored, among other things, the effects of changing age structures of the population or changes in nuptiality. The U.N. manual *Methodology of Measuring the Impact of Family Planning Programmes on Fertility* (United Nations, 1979) accordingly recommended "standardization" of rates as a first step.[2]

Similarly, the effect on fertility of higher ages of women at first marriage need to be taken into account. Even if the women who are waiting until their twenties to get married eventually catch up and have families of the same size as those who married earlier, the preliminary impact is to lower age-specific as well as crude fertility rates. For example, Hull, Hull, and Singarimbun (1977), conclude that 25 percent of the fertility decline observed in Indonesia was due to the rise in the age of first marriage. Unless the family planning program were somehow affecting the decisions of young people when to marry, fertility decline due to this source has to be subtracted from the total before the demographic impact of the program can be considered as a cause for the residual, unexplained part of the decline.

Substitution Effects

A more difficult problem has to do with estimating the degree to which a program substitutes public for private sources of modern contraceptives or modern for traditional contraceptives. Part of the service statistics of the program represent such substitution rather

[2]The manual is somewhat unclear about whether standardization is a first step or is, by itself, one of several roughly equivalent ways of estimating demographic impact. Some of the case studies presented in the companion volume appear to regard standardization as sufficient: Account for the fertility decline caused by changing age distribution and age at first marriage (sometimes marital separation and widowhood are included), then guess how much of the rest might be due to the program.

than a net increase in the total amount of contraception, and some of what should be attributed to the program is not measured, so the issue can be dropped. Since a public program may have had a "catalytic" effect on couples who do not get their contraceptive supplies from the program, by spreading the ideas of fertility regulation, some analysts have ignored the substitution effect--or progress in techniques of evaluation has been motivated by the research for a more sophisticated way of handling this problem.[3]

A preliminary approach to this problem is what the UN manual calls "trend analysis": graphing fertility rates against years, then fitting separate lines to the points for pre-program years and to the points for program years, to see if the rate of decline has increased. Informal as it may appear, this method has proved convincing, judging by the number of times it appears in the nontechnical literature.

For example, the U.S. Bureau of the Census *Country Demographic Profile* bases calculations of Indonesian fertility rates in the 1960s on the numbers of own children reported in the 1970 Census (U.S. Bureau of Census, 1979). It estimates a TFR of 5.66 over the period 1961-70, and a decline to 4.75-4.97 by 1976. The further decline in the late seventies has not been as great as was projected from these data, according to preliminary results of the 1980 census, but it is clear that fertility has declined and that the decline has been more rapid since 1970, when Indonesia's family planning program was fully operative, at least on Java and Bali, than before 1970. Unfortunately, even this method may demand too much data collection for some data-poor countries.

Even when good data on trends are available, this does not by itself solve the problem of separating out net program effects. A change in the rate of fertility decline is suggestive, nothing more. Fertility decline may have been building momentum for reasons independent of anything the family planning program did, and the program was just in place to take the credit. Lack of a sudden break in the trend shortly after the introduction of a program, conversely, would not be evidence of a lack of impact. Some traditional practices (abstinence

[3]Nortman et al. (1978) discuss this problem.

and prolonged breastfeeding) that had important effects on birth spacing may be on the wane. Valerie Hull's research suggests that this is just what is happening among poor women in Java (Hull, 1976). Likewise, the breastfeeding decline in Peninsular Malaysia has not been offset in its fertility impact by increased use of contraceptives, for some important subgroups of the population (DaVanzo and Haaga, 1982). If the effect of a family planning program is offsetting the effect of behavioral changes that would tend to *increase* fertility, this could be considered a demographic impact even in the absence of an acceleration of fertility decline.

At a micro level, various attempts have been made to estimate the potential fertility of program clients, based on an extrapolation of their previous fertility history, or based on matching records of clients with women of similar characteristics in the general population, whose subsequent fertility can then be compared. (Johnson et al., 1978, is a good example). The problems with this approach are that the two groups being compared are not necessarily similar: The program clients are a self-selected group, usually the more fecund women, with more children already born than the average, and with the most motivation to contracept. They may not be typical of the population as a whole. Nor can it be assumed that they will go on bearing children at the same rate as they have in the past, for they have access to traditional methods of contraception or abortion, and often to non-program sources of modern methods. Robert Potter and his associates have recently tried an ingenious method of estimating "fecundability functions" (what later fertility experience "would be") for particular subgroups of women, based on life-table analysis of birth intervals during which no contraception was used.[4] The empirical estimates were very sensitive to, among other things, the techniques used to handle the censored data points (Potter et al., 1979), but the approach is promising.

---

[4]"No contraception" is an ambiguous concept that is not easy to translate into questionnaire items, when rhythm, abstinence, breastfeeding, and/or spousal separation are practiced in part for contraceptive reasons.

Quasi-experiments

Other techniques use data at the level of the village, or service area of a clinic, or other geographic unit. One approach is similar to what Campbell and Stanley (1966) defined as "quasi-experiments" in social science research. Though it may be impossible to assign different areas randomly to experimental treatments or control groups, researchers could take advantage of the variation that exists among roughly similar units in the timing of introduction, intensity or type of family planning program operating in them. Multivariate statistical methods could be used to hold constant (account for variance in fertility outcomes caused by) some of the non-program factors. The independent effects of the program variables on fertility could then be estimated, subject to the completeness and correctness of the specification of the causal model used. Paul Schultz (1974) has used this method to good effect in a study of the family program on Taiwan.

But even if information on independent and dependent variables alike is available, there might not be enough variation in the timing of introduction on the type of program to produce such a natural experiment. For political reasons, it is often necessary to start a program everywhere at once (at least by declaration) to avoid giving the impression that some areas are being favored. Java and Bali, for example, seem to have been covered by the BKKBN (Indonesian family planning program) virtually all at once. The Serpong project in rural West Java, a long-term cooperative study by Dutch and Indonesian universities, was originally intended to include several control villages, but political and administrative considerations overruled the desire of the researchers for a well-designed study (Zwidberg, 1978). Or, if the timing of introduction is sufficiently staggered, there may be the problem that timing is itself related to an unobserved factor--the ability of the local administrators or headman to get resources for their villages--that is itself independently related to the health, fertility and other outcomes of interest.

Comparisons of geographic units that differ in the amount of program resources put into them can be made after the fact, even in the absence of a deliberately planned quasi-experiment. One study has

compared the recent history of fertility in Thailand's Chiang Mai province with the adjacent province of Chiang Rai, using data from pregnancy histories (Shevasunt, Hogan, and Thaithong, 1978). Chiang Mai was the scene of early family planning efforts funded by American philanthropy and, later, of strong efforts by the new NFPP. Chiang Rai was relatively untouched by promoters of family planning until recently. As might be expected, Chiang Mai had higher levels of usage of modern contraceptives than Chiang Rai. It has also had a more rapid decline in fertility rates than Chiang Rai, which suggests that the programs have had an impact. But the two provinces differ in ways that cannot be entirely accounted for within the samples drawn. Chiang Mai, for example, has at its center the second largest city of Thailand. Its people, even in rural areas, are more modern in many ways than those of Chiang Rai. The difference in fertility rates cannot be ascribed solely to the difference in NFPP resources.

This illustrates a general problem with the area-level approach: Even if data are available to control for household- or community-level differences in non-program factors affecting contraceptive use and fertility, these may be so correlated with program factors that their separate effect can not be untangled in a observational study. Planned, prospective studies, even experiments despite their difficulty, have some clear advantages over observational, after-the-fact studies. Their value may lie not so much in their application to global questions about the impact of a national-level programs, since such experiments are usually quite difficult to replicate on a large scale, but in their application to questions about the impact of incremental changes or particular variations of programs. Cuca and Pierce (1977) have prepared a comprehensive review of small-scale quasi-experiments in family planning programs.

Components Models

A third major approach works at the macro level, using simulation or "components" models to disaggregate sources of fertility decline or differences between current rates and those that would prevail with natural fertility in a population. First, estimates are obtained of the total amount of contraception (from all sources) that must have been

practiced in order to bring about an observed fertility decline. Next, program service statistics and assumed continuation and failure rates for different program-supplied contraceptives are applied to simulated cohorts of women. These are subsequently "aged," and the births averted by the use of program contraceptives are calculated, as is the gross impact of the program on fertility rates. Finally, the sensitivity of the estimate of the net impact of the program can be tested by varying assumptions about induced abortion, private sector contraceptive supply, breastfeeding, and other traditional contraceptive measures. Two roughly equivalent families of computer models are available: TABRAP/CONVERSE (Nortman et al., 1978) and PROJTARG (Teachman et al., 1978). Both were originally developed, not for retrospective evaluation, but for the setting of realistic targets for family planning programs.

Simulation models appear to be useful in program evaluation when at least some of the intermediate steps in the causal chain linking program output to demographic or health outcomes are known fairly precisely and are readily quantifiable. The sensitivity analyses performed by varying assumptions about the size and direction of other links in the causal chain are then useful, because the scope of uncertainty--the range of simulation results--is less after than it was before running the experiment. This is the case with simulation models in the epidemiology of infectious diseases, for example. The risk of contagion is related in nonlinear, but relatively well-known, ways to the proportion of the population that is susceptible to the disease. Also, the "use-effectiveness" of vaccines, like that of contraceptives, is usually known from field trials. If delivery statistics are available, useful estimates can be generated of the impact of immunization programs on disease incidence (Cvjetanovic et al., 1972). The situation is more complicated in family planning, however, because of the stronger possibility that program sources are substituting for others and the greater uncertainty about the effect of consumers' behavior on use-effectiveness.

In family planning, simulation models represent a sophisticated outgrowth of earlier methods that attempted to measure "couple-years of protection" or "births averted" by simple transformation of service

statistics.[5] In the more recent models, allowance is made for different continuation rates for contraceptive use across methods and across age groups of users. Also, the protection afforded by contraceptive use varies with the age of the user--secondary sterility, probabilities of widowhood, overlap of contraceptive use with postpartum anovulation, all are taken into account by the simulations. The result is that the gross estimates of program impact do not credit contraceptive use for averting births equally across all groups of women. These models are thus a decided advance over the methods that did not differentiate at all among users of contraceptive methods. They do require a good deal of data, which in practice often means the use of marital rates from old censuses, mortality rates from standard life tables, and the like.

Teachman et al. (1979), used these models with such data as were available to study four national family planning programs--the Thai and Indonesian programs, among others. The chapter on Thailand, written by Dennis Hogan, is worth following in some detail, because it shows how far it is possible to go with demographic data, program service statistics, and estimated rates of contraceptive usage *before* it is necessary to make assumptions about what fertility would have been in the absence of the program.

The first stage of Hogan's analysis involves calculating the "total number of woman-years of protection" required to bring the total fertility rate down from a presumed maximum of 8364 to what was actually observed in 1964 (before the NFPP), 1971 (when the NFPP was barely under way), and 1975 (the last year for which data were available). This part of the model "ages" cohorts of women, the size of the cohorts being taken from census data. Parameters that are used include age-specific mortality rates, proportions of women subject to primary sterility, proportions subject to secondary sterility, increasing with age, and proportions of women not in a sexual union, again by age.

The assumption that total fertility would reach 8.4 births per woman--the number 8364 cited above--is a weak point in this method. Theorists of "natural fertility" have calculated a wide range of values

---

[5]These methods are illustrated in the U.N. Manual (1978). See the criticisms of Reynolds (1972) and Schultz (1974).

for human populations ostensibly not regulating their fertility. The statement "The NFPP provided 61.6 percent of the total woman-years of contraception experienced by the Thai woman in 1975..." thus gives a misleading impression of precision.

The second stage uses NFPP service statistics (new acceptors by method per year) in the calculation of woman-years of protection obtained from the program. Besides service statistics, the data required at this stage include continuation rates by method (obtained by life-table techniques from surveys in Thailand).

Service statistics are inevitably of uncertain quality. There are potentially important differences in the way "acceptors" and "continuation" are defined, made operational, and measured. All that can be said is that evaluators have to be aware of the problem and seek data that are as comparable as possible across districts. Potter has criticized the use of life-table techniques to get continuation rates for methods other than the IUD: It is too simplifying to assume that use of the pill or the condom is "continuing" until evidence appears that it has stopped. Lastly, the numbers may be suspect if there is no reliable way of knowing if a woman has switched from one program-suppled method to another. Thus, the user of the PROJTARG or TABRAP models has to be satisfied that the effects of method-switching and varying effectiveness of use have been captured by the continuation rates input to the model.

At the final stage of the PROJTARG analysis, the women-years of contraception that the program provided is divided by the women-years of "total protection" to get the national proportion of protection to be ascribed to the program. If this is a high percentage (and if crude birth rates have, in fact, declined) the program is to be judged a success.

But what is put out of the model is still only a gross outcome measure, as the developers of PROJTARG were well aware. "Such a procedure assumes . . . that none of the protection provided by the NFPP would have been provided by other sources had the government program not existed." The question of what would have happened if there had been no active government policy, the authors judge, is "inherently unanswerable given the lack of any quasi-experimental Thai control group not exposed

to the NFPP . . ." (Teachman et al., p. 78). How, then, can some idea be formed of the net impact of the program?

Hogan approaches the problem in two ways. One is to assume that protection from private sources of both modern and traditional contraception would have increased at the same rate from 1971 to 1975 as it did from 1964 to 1971. Except for the refinement provided by the model's ability to handle changing proportions of women married, this is tantamount to the "trend analysis" discussed above. Any acceleration in fertility decline since the program started is ascribed to the program.

The second method, perhaps more convincing, is to make a conservative estimate of the proportion of program-supplied contraception and give reasons for thinking it conservative. Since even the highly refined service statistics are still producing only gross measures, Hogan performs a sensitivity analysis with different assumptions about the amount of substitution that could have occurred. He offers some reasons for thinking it no more than about half the equivalent NFPP contraception (denoted, as before, in women-years of protection of fecund women). Hogan observes that a high proportion of NFPP clients have been poor rural women, who would not otherwise have been able to afford efficient contraceptives, as shown by their low consumption of other modern medical supplies, etc. Abortions are prevalent in Thailand, and Hogan shows that the abortion rate would have had to double to fill the gap if the NFPP had not existed. Folk methods of contraception, inherently less effective, would have to have been practiced by far more women to make up for the women not using contraceptive pills supplied by the NFPP. The results of some rough calculations follow: Had the NFPP not existed, the observed decline of fertility by 1975 could have resulted if (1) over 769,000 of the 4,807,000 couples in the nation purchased their oral contraceptives; (2) about 1,120,000 couples practiced folk methods of contraception, or (3) approximately 380,000 couples resorted to abortion (Teachman et al., p. 65).

Conclusions (2) and (3) are somewhat overstated. Shifts in the age distribution of users of a particular non-program method, or increased effectiveness of use (averaged across all users), or shifts from less

effective to more effective methods within the class of non-program methods could have obviated some part of the need to recruit new "traditional contraceptors" in the absence of the program. For example, the effectiveness of abstinence or safe time methods would surely have been greater if the highly motivated couples who used the program had had to rely on them, instead of just the couples who did not know about or bother with the NFPP. Still, the analysis presented in Hogan's study of Thailand represents the current state of this art; it is hard to see how firmer inferences about program impact could have been drawn without information from experimental or quasi-experimental studies comparing similar groups with and without the program.

For many purposes, administrators may not care about the distinction between gross and net impacts. In allocating resources for program expansion, though, administrators could well care about the distinction and may need independent estimates of fertility decline, not just service statistics. This is the subject of the next section of this paper.

MARGINAL EFFECTIVENESS

Answering the question, "Did the program cause a drop in the birth rate?" does not exhaust the potential uses of evaluation research. Those who allocate resources for specific programs could use assistance in deciding where additional resources will do the most good. Assertions about the total effect of a program--averaged over all years, all components, and all areas served--may not address this concern.

For example, in Indonesia the city of Jakarta and the Outer Islands alike are held to be "under-served" by the BKKBN (Heiby, Ness, and Pillsbury, 1979). For Jakarta, this means that the percentage of women using program-supplied modern contraceptives is lower than in rural Java. For the Outer Islands, the problem may be one of access: Sulawesi and Borneo are not saturated with BKKBN distribution centers as are Java and Bali. How could the BKKBN establish priorities, ranking more intensive propaganda and recruiting in Jakarta versus more mobile clinics in Sulawesi? There are always considerations of equity, of course, both regional (if the government cannot start a program on one island and not extend it to the other islands) and across income groups

(poor people might be considered to need the program more than those better off). These would constrain program administrators, even if they had all the information they wanted about marginal effectiveness and costs of their activities in different regions. Insofar as the program goal is the simple one of reducing the nation's birth rate, though, the choice would involve a judgement on where the program's resources would have the greatest incremental effect on this rate. For example, it could be the case that fertility is declining in Jakarta anyway, because of non-program factors. Alternatively, it could be that there is a supply bottleneck in Jakarta, with a significant number of additional women who would limit their fertility if only the BKKBN would proselytize more actively. Residents of the Outer Islands may not be so well primed for the family planning program as were Java and Bali. A particularly important task in coming years will be to monitor the effects of the BKKBN as it moves to the Outer Islands. There is no reason to believe a priori that the same methods that worked so well in Java and Bali will work there. This is an opportunity for evaluators to be useful, instead of just punitive, from the point of view of program managers.

Berelson and Haveman (1980) raised the issue of measuring diminishing returns to program resources in a recent essay. They call for research that would "yield estimates of the relationship of expenditures on various options in various settings to the fertility reduction achieved, including the pattern of diminishing returns to incremental expenditures on each option." Such research is needed because "That knowledge does not exist at present" in usable form, but bits and pieces of it are scattered throughout the literature (Berelson and Haveman, 1980, p. 228).

The observation that there might be declining marginal returns to particular activities, and that this should affect managerial decisions, is not a new one in the field of family planning. Paul Schultz, in the study of Taiwan referred to above, was able to estimate production functions relating program inputs to that portion of the fertility reduction that varied with program expenditures and was not otherwise accounted for in his model. He found, as might be expected, that the effect of each local worker was greater in the early days when there

were few of them and declined over time as the program expanded and the island was more intensively covered. He also found that one sort of worker (the Village Health Education Nurse) had been more productive than another (the Pre-pregnancy Health Worker) at the proportions in which they were used by the Taiwanese program. Since the costs of training and paying them were not so different as their marginal products, Schultz is able to conclude that "the program would have had a substantially greater impact on birth rates in each year of the program if more VHENs and fewer PPHWs had been hired, holding constant total program expenditures (Schultz, 1974, p. 38). Again, the data available for Taiwan are unusually rich, but any contributions that program evaluators could make toward the assessment of the relative effectiveness of different parts of a program would clearly benefit policy planners.

Even if it is impossible to estimate a family planning "production function" reliably with the data at hand, evaluators could help by noting whether the characteristics of "non-acceptors" in the population change as the proportion of them decreases. P. M. Blaikie put the point this way in a review of the experience in India: "The response curve of total family planning acceptance to total number of family planning workers in any area . . . is probably curvilinear. . . . When the hard-core of non-adopters are left behind, it is time to divert programme resources to new areas" (Blaikie, 1972, p. 439). The conclusion does not follow necessarily, since consideration of substitution effects shows that "total family planning acceptance" is not the same thing as "fertility reduction achieved by the program," but the point is important. In the early stages, a program may be catering to a preexisting demand for contraceptives or spreading knowledge about modern methods to those who may never have heard of them. In later stages, the program's efforts are directed increasingly to those who, for various reasons, have heard about contraceptives but have not availed themselves of an opportunity to get them.

CONCLUSIONS

No amount of sophisticated manipulation of commonly available data can entirely put to rest an important question in the family planning debate: How much of a fertility decline would have happened anyway without free distribution of modern contraceptives, and how much is entirely attributable to the program? Studies of individual programs can increase the plausibility with which observed declines in fertility rates are ascribed to the efforts of family planning programs, but only an accumulation of rigorous studies in countries at different levels of administrative development and at different stages of demographic transition will provide convincing answers to the questions raised at the beginning of this Note. In the meantime, though, there will be pressure on donors and recipients alike to answer simple questions--Is this program working?--merely because funding for all development assistance will be even less lavish in coming years than it has been in the past. Planners and administrators can be usefully served by evaluations based on methods described above, leading, if not to a high degree of certitude, at least to some plausible guesses.

This plausibility will not come only from evaluations carried out at the level of the national program--the up-or-down, yes-or-no, continue-funding-or-don't evaluations. Neither will process evaluations, with their focus on what is needed for day-to-day operations and the meeting of targets, serve the purpose, however necessary these are for their own sake. In the case of well-established family planning programs, it may be impossible to introduce careful experimentation with control groups and random or almost-random assignment of subject to treatments.

What can be urged is a kind of intermediate-level evaluation, showing variations in the apparent impact of programs (as assessed by whatever of the methods outlined above are feasible) across regions, age, or ethnic groups, and over time. For the program manager, or for the budget analyst deciding whether to recommend that more money be sunk into family planning assistance, it may not be essential to know that a program has been, on the whole, a success. What is needed is a body of results (specific to one region or generalized as much as possible)

showing what seems to work, given what preconditions, and up to what point of diminishing returns. Such studies will have to entail assumptions about net, as opposed to gross, and marginal, as opposed to average, figures. Of the methods reviewed above, quasi-experimental studies, multivariate analyses of household survey data, and components models can all serve this purpose usefully, as long as the inevitable assumptions are made explicit.

REFERENCES

Berelson, B., and R. Haveman, "On Allocating Resources for Fertility Reduction in Developing Countries," *Population Studies*, Vol. 34, 1980.

Berelson, B., W. Mauldin, and S. Segal, "Population: Current Status and Policy Options," *Social Science and Medicine*, special issue on Health and Population in Developing Countries, Vol. 14C, No. 2, June 1980.

Blaikie, P. M., "Implications of Selective Feedback in Aspects of Family Planning Research for Policy Makers in India," *Population Studies*, Vol. 26, November 1972.

Campbell D., and J. Stanley, *Experimental and Quasi-experimental Designs for Research*, Rand McNally: Chicago, 1966.

Cuca R., and C. Pierce, *Experiments in Family Planning: Lessons from the Developing World*, Johns Hopkins University Press for the World Bank: Baltimore, 1977.

Cvjetanovic, B., B. Grab, K. Uemara, and B. Bytchenko, "Epidemiological Model of Tetanus and its Use in the Planning of Immunization Programs," *International Journal of Epidemiology*, Vol. 1, 1972, pp. 125-137.

Davis, K., "Population Policy: Will Current Programs Succeed?" *Science*, Vol. 158, November 19, 1967.

Heiby, J., G. Ness, and B. Pillsbury, *AID's Role in Indonesia Family Planning: A Case Study with General Lessons for Foreign Assistance*, Program Evaluation Report No. 2, U.S. Agency for International Development, Washington, D.C., 1979.

Hull, V., "The Positive Relation Between Economic Class and Family Size in Java: A Case Study of the Intermediate Variables Determining Fertility," Population Institute of Gadjah Mada University, Monograph Series No. 2, Yogyakarta, 1976.

Hull T., V. Hull, and M. Singarimbun, "Indonesia's Family Planning Story: Success and Challenge," *Population Bulletin*, Vol. 32, No. 6, November 1977.

Johnson, J. T., et al., "Impact of the Malaysian Family Planning Programme on Births: A comparison of Matched Acceptor and Non-Acceptor Birth Rates," *Population Studies*, Vol. 32, No. 2, July 1978, pp. 215-230.

Norton, D., R. Potter, S. Kirmeyer, and J. Bongaarts, *Birth Rates and Birth Control Practice*, Population Council, New York, 1978.

Potter, R, C. Mode, M. Soyka, and R. Avery, "Measuring Potential Fertility Through Null Segments: An Exploratory Analysis," *Social Biology*, Vol. 26, No. 4, Winter 1979.

Reynolds, J., "Evaluation of Family Planning Program Performance: A Critical Review," *Demography*, Vol. 9, No. 1, February 1972.

Schultz, T. P., *Fertility Determinants: A Theory, Evidence, and an Application to Policy Evaluation*, The Rand Corporation, R-1016-RF/AID, January 1974.

Shevasunt, S., D. Hogan, and K. Thaithong, "Fertility and Family Planning in Rural Northern Thailand," *Studies in Family Planning*, Vol. 9, August 1978.

Teachman, J., D. Hogan, and D. Bogue, "A Components Method for Measuring the Impact of a Family Planning Program on Birth Rates," *Demography*, Vol. 15, 1978, pp. 113-129.

Teachman, J., D. Bogue, J. Londono, and D. Hogan, *The Impact of Family Planning Programs on Fertility Rates: A Case Study of Four Nations*, Community and Family Study Center, University of Chicago, 1979.

United Nations, Department of International Economic and Social Affairs, *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*, Population Studies No. 66, Manual IX, ST/ESA/SER.A66, United Nations: New York, 1979.

U.S. Bureau of the Census, *Country Demographic Profiles: Indonesia*, ISP-DP-18, May 1979.

Wolfson, M., *Changing Approaches to Population Problems*, OECD, Paris, 1978.

Zwidberg, L. (ed.), *Family Planning in Rural West Java: The Serpong Project*, Universitas Indonesia--Leiden University, 1978.